Mastodon Theorem - 20 Years in the Making

P. D. Dragnev - Purdue University Fort Wayne (PFW)

Ratmino, July 20, 2019

*Jointly with Oleg Musin

Well-distributed points on the sphere

C is a point configuration (code) $C := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^2$:

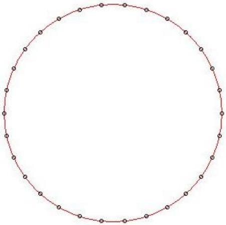
$$\mathbb{S}^2 = \{(x, y, z) \ : \ x^2 + y^2 + z^2 = 1\}$$

How do we distribute *well* the points of *C*?

2D problem - simple

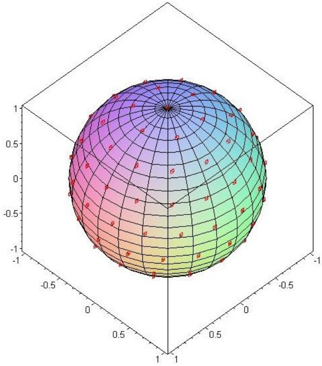
Answer: *Equaly spaced points*

Reason: *direction* and *order*.



Well-distributed points on the sphere

In 3D - no direction or order exists. Other methods and criteria are needed. To well-distribute means to **minimize potential energy.**



We distinguish:

Best packing points;

Fekete (Coulomb) points;

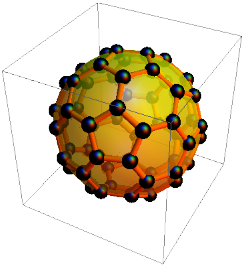
Logarithmic points;

Riesz points.

Why Minimize Potential Energy? Electrostatics:

Thomson Problem (1904) -
("plum pudding" model of an atom)

Find the (most) stable (ground state) energy configuration (**code**) of N classical electrons (Coulomb law) constrained to move on the sphere $\mathbb{S}^2$.



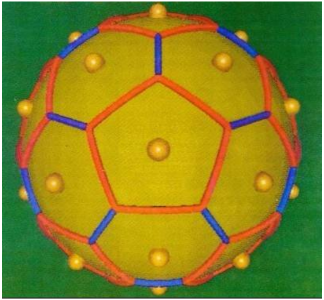
Generalized Thomson Problem ($1/r^s$ potentials and $\log(1/r)$)

A code $C := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^2$ that minimizes **Riesz *s*-energy**

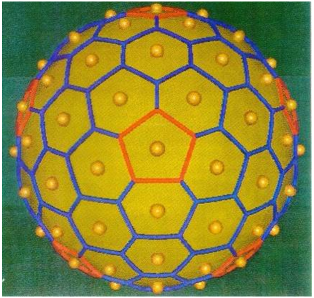
$$E_s(C) := \sum_{j \neq k} \frac{1}{|\mathbf{x}_j - \mathbf{x}_k|^s}, \quad s > 0, \quad E_{\log}(\omega_N) := \sum_{j \neq k} \log \frac{1}{|\mathbf{x}_j - \mathbf{x}_k|}$$

is called an **optimal *s*-energy code**.

Electrostatics and Electrons in Equilibrium



32 Electrons

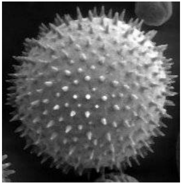


122 Electrons

Why Minimize Potential Energy? Coding Theory:

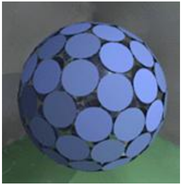
Tammes Problem (1930)

A Dutch botanist that studied modeling of the distribution of the orifices in pollen grain asked the following.



Tammes Problem (Best-Packing, $s = \infty$)

Place N points on the unit sphere so as to maximize the minimum distance between any pair of points.



Definition

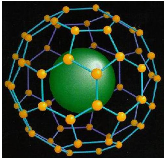
Codes that maximize the minimum distance are called **optimal (maximal) codes**. Hence our choice of terms.

More “Maximal Codes” from Biology

Why Minimize Potential Energy? Nanotechnology:

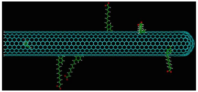
Fullerenes (1985) - (Buckyballs)

Vaporizing graphite, Curl, Kroto, Smalley, Heath, and O'Brian discovered C_{60} (Chemistry 1996 Nobel prize)



Nanotechnology - Nanowire (R. Smalley)

A giant fullerene molecule few nanometers in diameter, but hundreds of microns (and ultimately meters) in length, with electrical conductivity similar to copper's, thermal conductivity as high as diamond and tensile strength about 100 times higher than steel.



Fulerenes and Nanotechnology

The Discovery of the Buckyball C_{60} (1985)

- **Kroto (Sussex)** - studies chains of carbon atoms in intergalactic space;
- **Smalley, Curl (Rice)** - study clusters when vaporizing metal discs using laser-supersonic cluster beam apparatus;
- **Kroto** visits Rice, Sept 1, 1985, experiments start (graphite discs);
- Two graduate students, **Heath**, **O'Brien**, also involved;
- Mass spectrometer shows a large molecule, 720 amu, suggesting 60 carbon atoms;
- Paper submitted to NATURE on Sept 13, 1985;
- Accepted October 18, 1985;
- Published Nov. 14, 1985.

Richard Buckminster "Bucky" Fuller, 1895-1983

Montreal biosphere

Florida "Fullerene"

Epcot Center

Ancient "Fullerene"

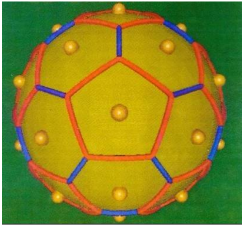
Under the lion paw

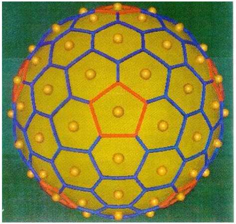
The "Mastodon" Fullerene Tzar

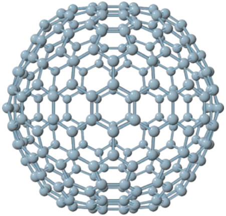
"Mastodon" Fullerene – C_{100} isolated

C_{100} **and Steven Stevenson**

Duality: 32/122 Electrons vs. C_{60}/C_{240} Buckyballs







Optimal s-energy codes on $\mathbb{S}^2$

Known optimal s-energy codes on $\mathbb{S}^2$

- $s = \log$, Smale's problem, logarithmic points (known for $N = 2 - 6,\ 12$);
- $s = 1$, Thomson Problem (known for $N = 2 - 6,\ 12$)
- $s = -1$, Fejes-Toth Problem (known for $N = 2 - 6,\ 12$)
- $s \to \infty$, Tammes Problem (known for $N = 1 - 12,\ 13, 14,\ 24$)

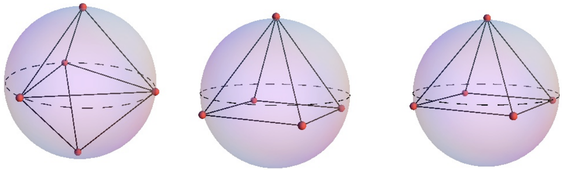
Limiting case - Best packing

For fixed N, any limit as $s \to \infty$ of optimal s-energy codes is an optimal (maximal) code.

Universally optimal codes

The codes with cardinality $N = 2, 3, 4, 6, 12$ are special (*sharp codes*) and minimize large class of potential energies. First "non-sharp" is $N = 5$ and very little is rigorously proven.

Optimal five point log and Riesz s-energy code on $\mathbb{S}^2$

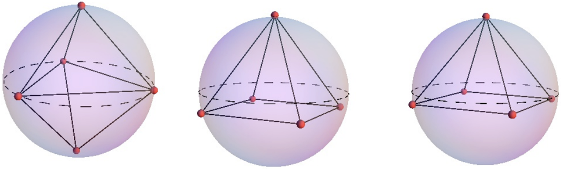


(a) (b) (c)

Figure: 'Optimal' 5-point codes on $\mathbb{S}^2$: (a) bipyramid BP, (b) optimal square-base pyramid SBP ($s = 1$) , (c) 'optimal' SBP ($s = 16$).

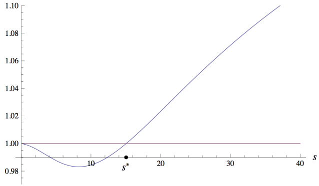
- $s = 0$: P. Dragnev, D. Legg, and D. Townsend, (2002) **(referred to by Ed Saff as "Mastodon" theorem)**;
- $s = -1$: X. Hou, J. Shao, (2011), computer-aided proof;
- $s = 1, 2$: R. E. Schwartz (2013), computer-aided proof;
- Bondarenko-Hardin-Saff (2014), As $s \to \infty$, any optimal s-energy codes of 5 limit is a square pyramid with base in the Equator;
- $0 < s < 15.04..$: R. E. Schwartz (2018).

Optimal five point log and Riesz s-energy code on $\mathbb{S}^2$



(a) (b) (c)

Figure: 'Optimal' 5-point code on $\mathbb{S}^2$: (a) bipyramid BP, (b) optimal square-base pyramid SBP ($s = 1$) , (c) 'optimal' SBP ($s = 16$).



Melnik et.el. 1977 $s^* = 15.04\ldots$?

Figure: 5 points energy ratio

"Mastodon" Theorem on $\mathbb{S}^3$ and $\mathbb{S}^4$ (Dragnev - 2016)

Definition

Two vertices $\mathbf{x}_i$ and $\mathbf{x}_j$ are called *mirror related* (we write $\mathbf{x}_i \sim \mathbf{x}_j$), if $|\mathbf{x}_i - \mathbf{x}_k| = |\mathbf{x}_j - \mathbf{x}_k|$, for every $k \neq i, j$.

Theorem (Characterization of $(d+3)$ Log-stationary configurations)

A log-stationary configuration is either (a) degenerate; (b) there exists a vertex with all edges stemming out being equal; or (c) every vertex is mirror related to another vertex.

Remark

Mirror relation is equivalence relation and an equivalence class forms a regular simplex in the spanning affine hyperspace.

Theorem (Dragnev - 2016)

The $(d+3)$-Log-optimal configuration in $\mathbb{S}^1, \mathbb{S}^2, \mathbb{S}^3, \mathbb{S}^4$, is two orthogonal simplexes of type $\{2,2\}, \{2,3\}, \{3,3\}, \{3,4\}$ respectively.

“Mastodon” Theorem on $\mathbb{S}^{d-1}$ (Musin, D. - 2018)

Theorem (Main Theorem 1)

Up to orthogonal transform, every relative minimum of the logarithmic energy $E_{\log}(X)$ of $d+2$ points on $\mathbb{S}^{d-1}$ consists of two regular simplexes of cardinality $m \geq n > 1$, $m+n=d+2$, such that these simplexes are orthogonal to each other. The global minimum occurs when $m=n$ if d is even and $m=n+1$ otherwise.

Stationary Configurations of $d+2$ points on $\mathbb{S}^{d-1}$

Theorem (Main Theorem 2)

Let $N = d+2$ and $X = \{x_1, \ldots, x_N\}$ be a non-degenerate stationary logarithmic configuration on $\mathbb{S}^{d-1}$. Suppose there is no point $x \in X$ that is equidistant to all other points in X. Then X can be split into two sets such that these sets are vertices of two regular orthogonal simplexes with the centers of mass in the center of $\mathbb{S}^{d-1}$.

Remark

Strengthens 2016 Characterization theorem significantly.

Stationary Configurations of $d+2$ points on $\mathbb{S}^{d-1}$

Given potential interaction function $h : [-1,1] \to \overline{\mathbb{R}}$ *h-energy* is

$$E_h(X) := \sum_{1 \leq i \neq j \leq N} h(x_i \cdot x_j).$$

Theorem (Degenerate Case)

Let X be a degenerate configuration, $N \geq d+2$, and $h : [-1,1] \to \mathbb{R}$ be a strictly convex potential function. Then there exists a continuous perturbation that decreases the h-energy $E_h(X)$.

Theorem (Equidistant case)

A non-degenerate stationary log-energy configuration of type $\{1, 1, \ldots, k, l\}$, where $1 + 1 + \cdots + k + l = d + 2$ is a saddle point. Moreover, there is a continuous perturbation that decreases the logarithmic energy of the $\{1, k, l\}$ part of the configuration to either $\{k+1, l\}$ or $\{k, l+1\}$. Sequence of such perturbations leads to relative minima as described in Main Theorem.

Radon's theorem

Theorem

Let $N = d + 2$ and $X = \{x_1, \dots, x_N\}$ be a set of points in $\mathbb{R}^d$. Then X can be partitioned into two disjoint sets whose convex hulls intersect.

Proof.

There exists a set of multipliers $a_1, ..., a_{d+2}$, not all of which are zero, solving the system of linear equations

$$\sum_{i=1}^{d+2} a_i x_i = 0, \qquad \sum_{i=1}^{d+2} a_i = 0.$$

Let $I := \{i \mid a_i \geq 0\}$ and $J := \{i \mid a_i < 0\}$. Then

$$p = \sum_{i \in I} \frac{a_i}{A} x_i = \sum_{i \in J} \frac{-a_i}{A} x_i, \qquad A = \sum_{i \in I} a_i = -\sum_{i \in J} a_i.$$

□

Auxiliary Results

Using Lagrange Multipliers method to *logarithmic energy*

$$E_{Log}(X) := -\frac{1}{2} \sum_{1 \leq i \neq j \leq N} \log(x_i \cdot x_i - 2x_i \cdot x_j + x_j \cdot x_j),$$

and differentiating yields

$$\sum_{j \neq i} \frac{x_i - x_j}{r_{i,j}} = \lambda_i x_i \quad i = 1, \ldots, N, \text{ where } r_{ij} := 1 - x_i \cdot x_j.$$

Taking inner product of both sides with x_i one obtains $\lambda_i = N - 1$, or

$$\sum_{j \neq i} \frac{x_i - x_j}{r_{i,j}} = (N-1)\, x_i, \quad i = 1, \ldots, N. \tag{1}$$

Summing (1) implies that the centroid lies at the origin, and hence

$$\sum_{j} r_{ij} = N, \quad i = 1, \ldots, N. \tag{2}$$

Auxiliary Results - Rank Lemma

Let

$$B=(b_{ij}), \quad b_{ij}:=\frac{1}{r_{ij}}, \quad b_{ii}:=N-1-\sum_{j\neq i} b_{ij},$$

$$A=(a_{ij}), \text{ where } a_{ij}:=c-b_{ij}, \quad c:=\frac{N-1}{N}.$$

Lemma

Let $X=\{x_1,\dots,x_N\}$ be a stationary logarithmic configuration on $\mathbb{S}^{d-1}$ that is non-degenerate ($span(X)=\mathbb{R}^d$). Then

$$\operatorname{rank}(A)\leq N-d-1, \qquad \sum_{j=1}^{N} a_{ij}=0, \quad i=1,\dots,N.$$

If $N=d+2$, then $\operatorname{rank}(A)=1$.

Proof of the Rank Lemma

Let $\mathrm{X} := [x_1, \ldots, x_N]^T$. The force equations (1) and (2) imply that

$$\sum_{j=1}^{N} b_{ij} x_j = 0, \quad \sum_{j=1}^{N} b_{ij} = N - 1.$$

In other words, $B\mathrm{X} = 0$ and $B\mathbf{1} = (N-1)\mathbf{1}$, where $\mathbf{1}$ denotes the N-dimensional column-vector of ones. As X is non-degenerate, we have rank $\mathrm{X} = d$. Therefore, the column-vectors of X are linearly independent. As $\mathbf{1}$ is eigenvector of B with an eigenvalue of $N-1$ it is linearly independent to the columns of X (eigenvectors with eigenvalue 0). The lemma follows from the rank-nullity theorem applied to $A[\mathrm{X}, \mathbf{1}] = 0$.

Auxiliary Results - $N = d + 2$

The following lemma elaborates on the case when $N = d + 2$.

Lemma

Let $N = d + 2$ and $X = \{x_1, \dots, x_N\}$ be a non-degenerate stationary logarithmic configuration on $\mathbb{S}^{d-1}$. Without loss of generality we may assume that $a_{1i} \geq 0$ for $i = 1, \dots k$ and $a_{1i} < 0$ for $i = k + 1, \dots N$. Let

$$a_i = \sqrt{a_{ii}},\ i = 1, \dots k;\ \ a_i = -\sqrt{a_{ii}},\ i = k + 1, \dots N.$$

Then

$$a_{ij} = a_i\, a_j, \quad a_1 + \dots + a_N = 0,$$

$$c - a_i a_j \geq \frac{1}{2},\ \ \text{for all}\ \ i \neq j,$$

$$\sum_{j \neq i} \frac{1}{c - a_i a_j} = N,\ i = 1, \dots, N. \tag{3}$$

Auxiliary Results - Supplemental Theorem

If $a_i = 0$ then the i-th row and i-th column in A are zero x_i is equidistant to all other points x_j. So, $a_i \neq 0$ for all $i = 1, \ldots, N$.

Theorem (Supplemental)

Let $a_1, \ldots, a_N$ be real numbers that satisfy the following assumptions

$$a_1 \geq \ldots \geq a_k > 0 > a_{k+1} \geq \ldots \geq a_N, \quad a_1 + \ldots, + a_N = 0,$$

$$\sum_{j \neq i} \frac{1}{c - a_i a_j} = N, \; i = 1, \ldots, N, \quad c - a_i a_j > 0, \; \textit{for all} \; i \neq j,$$

where $c := \frac{N-1}{N}$. Then

$$a_1 = ... = a_k, \quad a_{k+1} = ... = a_N.$$

Auxiliary Results - Technical Lemma

Lemma (Technical)

Suppose $a_1, \ldots, a_N$ *are as in Supplemental Theorem. Then for all* $i = 1, \ldots, N$ *we have*

$$T_i := \sum_{j=1}^{N} \frac{c - a_j^2}{c - a_i a_j} = N - 1. \quad (4)$$

and

$$|a_i| < \sqrt{c}, \quad i = 1, ..., N. \quad (5)$$

Proof of Supplemental Theorem

Let

$$F(t) := \sum_{j=1}^{N} \frac{c - a_j^2}{c - ta_j}.$$

Then Technical Lemma implies that for all $i = 1, \ldots, N$

$$F(a_i) = N - 1. \tag{6}$$

Since

$$F''(t) = 2 \sum_{j} \frac{\left(c - a_j^2\right) a_j^2}{(c - ta_j)^3},$$

by Technical Lemma again we have $F''(t) > 0$ for $t \in (-\sqrt{c}, \sqrt{c})$. Hence $F(t)$ is a convex function in this interval. Therefore, the equation $F(t) = N - 1$ has at most two solutions. By assumptions we have $a_i > 0$ for $i = 1, \ldots, k$ and $a_i < 0$, for $i = k + 1, \ldots, N$. Thus, (6) yields that all positive a_i are equal and all negative a_i are equal too. $\Box$

Inequality 1

Lemma (H1)

Let $A = (a_{ij})$ *be an* $m \times m$ *matrix,* $m \geq 3$*, such that*
(a) $a_{ii} = 0, \quad i = 1, \ldots, m;$
(b) $\sum_{j=1}^{m} a_{ij} = 0.$
Then the following inequality holds

$$\sum_{1 \leq i < j \leq m} (a_{ij} + a_{ji})^2 \geq \frac{1}{m-2} \sum_{j=1}^{m} x_j^2, \quad \text{where} \quad x_j := \sum_{i=1}^{m} a_{ij}. \qquad (7)$$

Proof of Inequality 1: part 1

For all $i, j = 1, \ldots, m$ define

$$\beta_{ij} := \frac{1}{m^2 - 2m} x_i + \frac{m-1}{m^2 - 2m} x_j, \quad i \neq j, \quad \text{and} \quad \beta_{ii} = 0.$$

Since $\sum_{j=1}^{m} x_j = 0$, we have $\sum_{j=1}^{m} \beta_{ij} = 0$ and $\sum_{i=1}^{m} \beta_{ij} = x_j$, i.e.

$$\sum_{j=1}^{m} \beta_{ij} = \sum_{j=1}^{m} a_{ij} \quad \text{and} \quad \sum_{i=1}^{m} \beta_{ij} = \sum_{i=1}^{m} a_{ij}.$$

Let $\widetilde{a}_{ij} := a_{ij} - \beta_{ij}$. Then

$$\sum_{i} \widetilde{a}_{ij} = \sum_{j} \widetilde{a}_{ij} = 0.$$

Proof of Inequality 1: part 2

Consider $t_{ij} := a_{ij} + a_{ji} = w_{ij} + \beta_{ij} + \beta_{ji}$, where $w_{ij} = \widetilde{a}_{ij} + \widetilde{a}_{ji}$. Then $t_{ij} = w_{ij} + \frac{x_i}{m-2} + \frac{x_j}{m-2}$, $i \neq j$, where $\sum_i w_{ij} = \sum_j w_{ij} = 0$ (observe that $t_{ii} = 0$). Then

$$\sum_{i<j} t_{ij}^2 = \sum_{i<j} \left(w_{ij} + \frac{x_i}{m-2} + \frac{x_j}{m-2} \right)^2 = \sum_{i<j} w_{ij}^2 + \frac{1}{m-2} \sum_{i=1}^{m} x_i^2,$$

which implies (7).

Inequality 2

Lemma (H2)

Given an $m \times n$ matrix $F = (f_{ij})$ and an $n \times m$ matrix $G = (g_{ij})$ such that $\sum_{j=1}^{n} f_{ij} = 0$ for all $i = 1, \ldots, m$ and $\sum_{j=1}^{m} g_{ij} = 0$ for all $i = 1, \ldots, n$. Then we have

$$\sum_{i=1}^{n} \sum_{j=1}^{m} (f_{ij} + g_{ji})^2 \geq \frac{1}{m} \sum_{j=1}^{n} y_j^2 + \frac{1}{n} \sum_{i=1}^{m} z_i^2,$$

$$y_j := \sum_{i=1}^{m} f_{ij}, \quad z_i := \sum_{j=1}^{n} g_{ji}.$$

Proof of Inequality 2

Let

$$\widetilde{f}_{ij} := f_{ij} - \frac{y_j}{m} \quad \text{and} \quad \widetilde{g}_{ij} := g_{ij} - \frac{z_i}{n}.$$

Since $\sum_j y_j = \sum_i z_i = 0$, we have $\sum_{i,j}(\widetilde{f}_{ij} + \widetilde{g}_{ji}) = 0$. Let $t_{ij} := \widetilde{f}_{ij} + \widetilde{g}_{ji}$. Observe that

$$\sum_{i=1}^{m} t_{ij} = \sum_{j=1}^{n} t_{ij} = 0.$$

From

$$f_{ij} + g_{ji} = \frac{y_j}{m} + \frac{z_i}{n} + t_{ij}.$$

one derives that

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(f_{ij}+g_{ji})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{y_j}{m} + \frac{z_i}{n} + t_{ij}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} t_{ij}^2 + \frac{1}{m}\sum_{j=1}^{n} y_j^2 + \frac{1}{n}\sum_{i=1}^{m} z_i^2,$$

which completes the proof.

Proofs of Degenerate, Equidistant, and Relative Minima cases

Even more complex and involved :-(

Degenerate case

Theorem (Degenerate Case)

Let X be a degenerate configuration, $N \geq d+2$, and $h : [-1,1] \to \mathbb{R}$ be a strictly convex potential function. Then there exists a continuous perturbation that decreases the h-energy $E_h(X)$.

Proof.

$$x_1 = (r, \sqrt{1-r^2}, 0, \dots, 0), x_2 = (r, -\sqrt{1-r^2}, 0, \dots, 0)$$

$$x_j = (c_{j1}, c_{j2}, c_{j3}, \dots, 0),\ j = 3, \dots, N,$$

where $c_{32} \neq 0$. Preturb to $\widetilde{X}$

$$\tilde{x}_1 = (r, \sqrt{1-r^2}\cos\theta, 0, \dots, \sqrt{1-r^2}\sin\theta),$$

$$\tilde{x}_2 = (r, -\sqrt{1-r^2}\cos\theta, 0, \dots, -\sqrt{1-r^2}\sin\theta).$$

Then $E_h(X) > E_h(\widetilde{X})$ □

Equidistant case

Theorem (Equidistant case)

A non-degenerate stationary log-energy configuration of type $\{1, 1, \dots, k, l\}$, where $1 + 1 + \cdots + k + l = d + 2$ is a saddle point. Moreover, there is a continuous perturbation that decreases the logarithmic energy of the $\{1, k, l\}$ part of the configuration to either $\{k + 1, l\}$ or $\{k, l + 1\}$. Sequence of such perturbations leads to relative minima as described in Main Theorem.

Equidistant case proof

Proof.

Let $X = \{1, k, \ell\}$ with $x_N \cdot x_i = -1/(N-1)$. Denote $x_i = (y_i, \frac{-1}{N-1})$, $z_i := (N-1)y_i/\sqrt{N(N-2)}$, $z_i \in \mathbb{S}^{d-2}$ satisfies force equation.

$$Y := \{(\sqrt{1 - 1/(k+m)^2}\, y_i, 0_{m-1}, -1/(k+m))\},$$

$$Z := \{(0_{k-1}, \sqrt{1 - 1/(k+m)^2}\, z_j, -1/(k+m))\}$$

Perturb to

$$\widetilde{Y}_t = \left\{ \left(\sqrt{1 - (mt + 1/(k+m))^2}\, y_i, 0_{m-1}, -1/(k+m) - mt \right) \right\}_{i=1}^{k}$$

$$\widetilde{Z}_t = \left\{ \left(0_{k-1}, \sqrt{1 - (kt - 1/(k+m))^2}\, z_j, -1/(k+m) + kt \right) \right\}_{j=1}^{m}.$$

Then $E_h(\widetilde{X}_t)$ has local max at $t = 0$ and decreases to $\{k, \ell + 1\}$ or $\{k+1, \ell\}$. □

Relative minima case

Theorem (Equidistant case)

Let $X = \{k, \ell\}$ a configuration of two orthogonal simplexes X_k and X_ℓ. Any perturbation will increase the energy locally.

THANK YOU!